

**Reconstruction of concentrations from
particle distributions in
high dimensional spaces**

Treball realitzat per:

Esteban Torres Centeno

Dirigit per:

Daniel Fernàndez Garcia

Llicenciatura en:

Enginyeria Geològica

Barcelona, 20 de Setembre de 2016

Departament d'Enginyeria Civil i Ambiental

TREBALL FINAL DE CARRERA

Agraïments

Primer de tot, agrair al meu tutor Daniel Fernández Garcia per l'ajut i la dedicació en aquest treball, per l'après i gaudir treballant amb ell.

En segon lloc, donar les gràcies als amics que han suportat els mals humors de les hores de programació, i també les bones estones de revisió de faltes d'ortografia.

De la mateixa manera, donar les gràcies a la meva família per l'empenta en aquests mesos, sobretot als meus pares Esteban i Trinitat, que han estat presents en tot moment.

Finalment, amb especial agraïment a l'Eduarne, per mai deixar de creure en mi.

Resum

Els *Kernel Density Estimators* són una òptima resolució per a problemes concentracions de contaminants en aigües subterrànies. La seva particularitat, el tractament de les dades a partir de posicions i la reconstrucció en concentracions, és un dels atractius d'aquest mètodes degut al bon comportament respecte problemes d'advecció-dispersió. Amb la base científica dels mètodes lagrangians com el *Random Walk Particle Tracking*, donen l'ajut per donar la continuïtat necessària que no obteníem amb altres resolucions com models de caixes continus.

L'objectiu del treball és tractar, estimar i comparar, dades de partícules aleatòries que emulin diferents contaminants transportats per un flux subterrani, on els mètodes de recollida de dades són limitats, degut a la barrera física que trobem. Per això, ens basem en models de reconstrucció de partícules on l'aleatorietat és fonamental per abraçar el màxim de comportament de contaminants, sense deixar de banda l'elecció del mètode de resolució i dels paràmetres adients d'aquest. El repte principal d'aquest estudi serà l'obtenció d'un mètode de resolució de concentracions a partir de posicions de partícules òptim, on s'exposarà i es tractarà els paràmetres que el caracteritzen.

Per obtenir una visió global de la hidrologia subterrània, el **capítol 2**, resumeix l'estat de l'art d'aquesta ciència, fent èmfasi en les equacions que tracten el medi.

En els **capítols 3 i 4** es presenta el transport de contaminants, els principals processos que el caracteritzen i la materialització física d'aquests mitjançant les equacions pròpies que s'escauen. S'aprofundeix en els mètodes lagrangians, concretament en el *Random Walk Particle Tracking*, mètode emprat en la resolució de densitats de partícules a partir de posicions de partícules aleatòries. Es tractaran els *Kernel Density Estimators* com a part fonamental de la resolució del mètode i del seu coeficient de *Bandwidth* com a clau d'èxit en la semblança de l'aproximació a la realitat.

Al **capítol 5** es mostra el codi MATLAB utilitzat en el tractament de dades de partícules aleatòries.

Finalment, en el **capítols 6 i 7** es discuteixen els resultats i es donen les conclusions locals i globals del mètode escollit en el projecte.

Abstract

The Kernel Density Estimators are an optimal solution for solving the pollutant concentrations problems in groundwater. The main feature of this solution is the data treatment based on taking and transforming positions into concentrations because it works efficiently against advection-dispersion problems. Using The *Random Walk Particle Tracking* as a base for this solution, you can get the necessary continuity that you wouldn't get using other methods like the continuous box methods.

The aim of this research is to try, estimate and compare random particles' data while they are flowing in groundwater, where the methods to collect samples are physically limited. This is the reason why we base our research on the particles' reconstruction models, where randomness is fundamental to cover as many pollutant performances as possible but keeping in mind the method chosen before. The main challenge of this research will be computing an optimal reconstruction of concentrations method from particles where we will show the factors that characterize it.

In **chapter 2**, you will be able to learn the most important features about groundwater, emphasizing the main equations.

Chapters 3 and 4 show the pollutant transport, the main processes that characterize it and the physical performances. Furthermore, in these chapters we will go into detail about the lagrangian method, specifically the *Random Walk Particle Tracking*, which will be used to know the particles' density coming from random particles. Kernel Density Estimators will be used as the key method to estimate density and the Bandwidth coefficient as the main parameter.

Chapter 5 offers the MATLAB code to estimate random particle densities.

Finally, **chapters 6 and 7** will show the results and the local and global conclusions of the method chosen for this research.

Sumari

Agraïments	3
Resum	4
Abstract	5
Sumari	6
Índex de figures	8
Índex de taules	10
Glossari	11
Prefaci	12
1. Introducció	14
1.1 Objectiu del projecte	14
1.2 Abast del projecte	14
2. Conceptes generals d'hidrologia subterrània	16
2.1 Hidrologia subterrània	16
2.1.1 Cicle de l'aigua subterrània	17
2.1.2 Formacions geològiques subterrànies	18
2.2 Equació del flux en medi porós	19
2.2.1 Llei de Darcy	19
2.2.2 Principi de conservació de la massa en un volum de control.	21
2.2.3 Equació del flux	21
3. Introducció al transport de contaminants	25
3.1 Mecànica del transport en aigües subterrànies	25
3.1.1 Processos de contaminants reactius	25
3.1.2 Processos de contaminants no reactius	26
3.1.3 Equació Advecció- Dispersió	31
3.2 Mètodes numèrics aplicats al transport de contaminants	31
3.2.1 Mètodes Eulerians	31
3.2.2 Mètodes Lagrangians	32
3.2.3 Mètodes Mixtes (Eulerià-Lagrangià)	32

4. Particle Tracking Methods (PTMs)	34
4.1 Random Walk Particle Tracking (RWPT).....	34
4.2 Kernel Density Estimators (KDE).....	37
4.3 Bandwidth (h)	40
4.3.1 Rules of Thumb	40
4.3.2 Least Squares Cross-Validation.....	41
4.3.3 Solve-the-Equation Plug-in Approach	42
5. MATLAB: Resolució de densitats de partícules.....	44
6. Resultats.....	49
6.1 Distribució Normal.....	49
6.1.1 Distribució Normal: Número de punts	50
6.1.2 Distribució Normal: Bandwidth.....	51
6.1.3 Distribució Normal: Error.....	52
6.2 Distribució Lognormal.....	52
6.2.1 Distribució Lognormal: Número de punts	53
6.2.2 Distribució lognormal: Bandwidth.....	54
6.2.3 Distribució Lognormal: Error	55
6.3 Distribució Beta	55
6.3.1 Distribució Beta: Número de punts	56
6.3.2 Distribució Beta: Bandwidth.....	57
6.3.3 Distribució Beta: Error.....	58
6.4 Distribució Bimodal.....	58
6.4.1 Distribució Bimodal: Número de punts	59
6.4.2 Distribució Bimodal: Bandwidth.....	60
6.4.3 Distribució Bimodal: Error	61
7. Conclusions	63
Bibliografia.....	64

Índex de figures

Figura 2.1 Cicle de l'aigua (Met Office UK Government, 2015).....	16
Figura 2.2 Cicle de l'aigua ampliat. Aigua subterrània emmagatzemada en formacions geològiques, infiltrada per zones de carrega (infiltració per pluges i rius) i zones de descarrega (pous, llacs i mar). (Alfonso Rivera, 2008).....	17
Figura 2.3. Permeàmetre utilitzat per Henry Darcy al seu experiment. (F. Javier Sánchez San Román, Juliol 2014).....	19
Figura 3.1. Transport únicament per advecció (F. Javier Sánchez San Román, Juny 2012).	26
Figura 3.2. Transport combinatiu de difusió i advecció (F. Javier Sánchez San Román, Juny 2012).	27
Figura 3.3. Dispersió mecànica en un medi porós. (F. Javier Sánchez San Román, Juny 2012).	29
Figura 6.1. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució normal. (Extreta de MATLAB)	50
Figura 6.2. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució normal. (Extreta de MATLAB).	51
Figura 6.3. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució normal, (Extreta de MATLAB).	52
Figura 6.4. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució lognormal. (Extreta de MATLAB).	53
Figura 6.5. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució lognormal. (Extreta de MATLAB).	54
Figura 6.6.. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució lognormal, (Extreta de MATLAB).	55
Figura 6.7. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució Beta ($A=1$, $B=3$). (Extreta de MATLAB).....	56
Figura 6.8. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució beta ($A=1$, $B=3$). (Extreta de MATLAB).....	57
Figura 6.9. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució Beta ($A=1$, $B=3$), (Extreta de MATLAB).	58
Figura 6.10. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució Bimodal. (Extreta de MATLAB)	59
Figura 6.11. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució bimodal. (Extreta de MATLAB).	60

Figura 6.12. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució bimodal, (Extreta de MATLAB)..... 61

Índex de taules

Taula 4.1 Exemples de la funció del Kernel (Silverman, 1986)	38
--	----

Glossari

ADE: Advection-Dispersion Equation

ETC: Equació Transport de Contaminants

PTMs: Particle Tracking Methods

BTCs: Breakthrough Curves

RWPT: Random Walk Particle Tracking

KDE: Kernel Density Estimators

MISE: Mean Integrated Squared Error

AMISE: Asymptotic Mean Integrated Squared Error

BW: Bandwidth

Prefaci

La contaminació d'aigües subterrànies és una de les problemàtiques que ens trobem a finals del S.XX i principis del S.XXI. L'explotació de recursos vitals com és l'aigua dolça en medis subterranis ha fet sorgir el nivell de contaminació d'alguns medis amb la conseqüència de la inutilització del consum d'aquesta aigua.

Quan un contaminant entra en contacte amb aigua subterrània, aquest adquireix una relació amb el flux, on presenta diferents processos físics i químics. El procés que pateix el contaminant és complex i difícil de modelar, no obstant existeixen mètodes especialitzats en aquests diversos comportaments que garanteixen la similitud amb la realitat.

La possibilitat de modelar aquest procés, veure com es desenvolupa és una eina molt interessant, que ens mostra no només a la categorització dels medis soterrats, sinó que ajuda al tractament i a la posterior recuperació del recurs.

El *Random Walk Particle Tracking* és un mètode òptim per resoldre problemes de transport dominats principalment per l'advecció. Tractem l'advecció com l'arrossegament del contaminant pel flux d'aigua, principal moviment que pateix tot contaminant. La trajectòria que presenten i el desplaçament de les partícules és important, tant com la concentració de contaminant que hi ha en el medi, degut als diferents índex mínims de concentracions de contaminants en aigües subterrànies que depenen de la perillositats d'aquests.

Amb l'ajut dels *Kernel Density Estimators*, aquest mètode obté una gran utilitat ja que aporta la continuïtat necessària en el tractament de transport de contaminants, on observem en cada moment una relació densitat-posició, més que útil per la modelització final.

Capítol 1:

Introducció

1. Introducció

1.1 Objectiu del projecte

Els objectius principals d'aquest projecte són:

- Definició del model de reconstrucció de partícules a partir de concentracions en grans espais dimensionals.
- Caracteritzar els paràmetres i variables del model de reconstrucció de partícules.
- Avaluació del model, amb exposició de diferents paràmetres.

1.2 Abast del projecte

El projecte abasta els punts següents:

- L'anàlisi del model de reconstrucció.
- Optimització dels paràmetres del model.
- Desenvolupament del model de reconstrucció en 1D.

Capítol 2:

Conceptes generals d'hidrologia subterrània

2. Conceptes generals d'hidrologia subterrània

L'aigua és fonamental per a la vida de l'ésser humà. Si ens aturem a pensar en tot els usos que l'hi donem avui dia al S.XXI ens fariem creus de la gran quantitat d'informació, processos i estudis en els que ella es present. No només a aquest darrer segle sinó durant l'història de la humanitat l'aigua ha estat gran aliada i alhora enemiga també, dels habitants d'aquest planeta. Alimentar-se i sobreviure ha estat base de l'existència i l'aigua no ens ha deixat de banda amb rius, torrents, mars, pluges, nevades, pous i llacs els quals formen part del cicle de l'aigua (*Figura 2.1*).

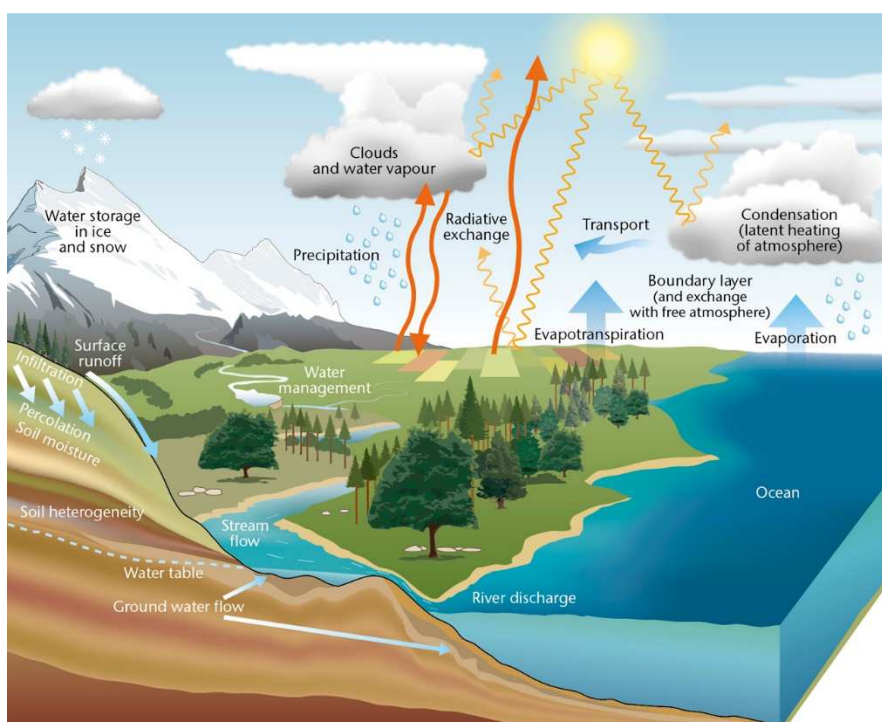


Figura 2.1 Cicle de l'aigua (Met Office UK Government, 2015)

Part d'aquest cicle de l'aigua, és aquella que no és visible des de la superfície terrestre, la que hem extret durant mil·lennis de pous o hem pescat en ella en llacs, l'aigua subterrània.

2.1 Hidrologia subterrània

La hidrologia subterrània és la ciència la qual compren el desenvolupament i control de les aigües subterrànies en benefici de la humanitat. El seu estudi abraça la determinació de les propietats i característiques de les aigües subterrànies, les quals saturen les formacions geològiques, aplicant els principis hidràulics bàsics.

2.1.1 Cicle de l'aigua subterrània

El paper del cicle de l'aigua en la hidrologia subterrània és fonamental, degut a l'entrada d'aigua al sistema subterrani amb pluja o neu que s'infiltra al subsòl, donant sortida per llacs, pous o desembocant al mar.

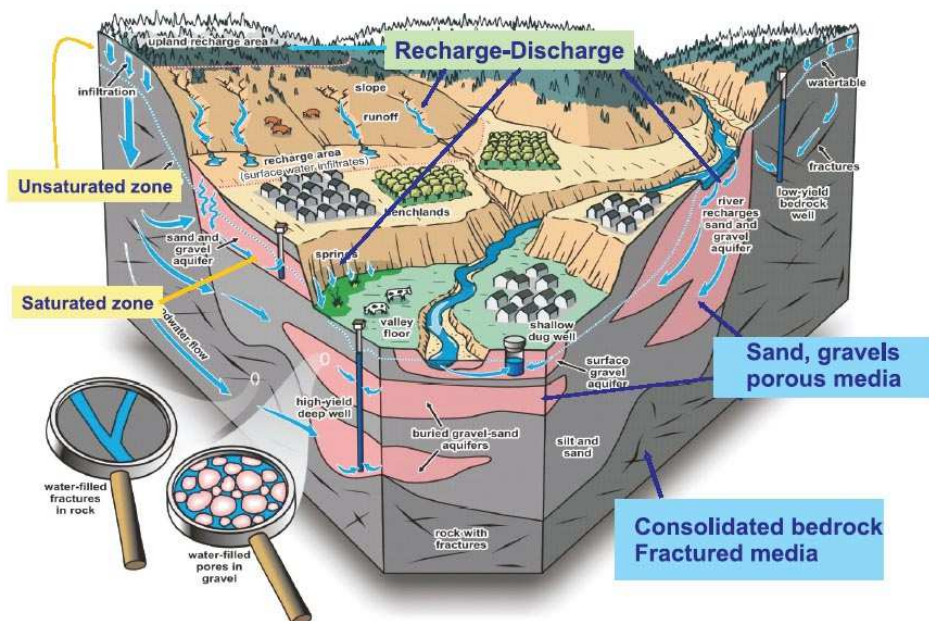


Figura 2.2 Cicle de l'aigua ampli. Aigua subterrània emmagatzemada en formacions geològiques, infiltrada per zones de carrega (infiltració per pluges i rius) i zones de descarrega (pous, llacs i mar). (Alfonso Rivera, 2008)

A la *Figura 2.2* veiem com les zones de càrrega, infiltren l'aigua cap a l'interior de la terra donant a conèixer tres zones del sistema subterrani:

- Zona no saturada: l'aigua que s'infiltra al sistema per aquesta zona edàfica on es conduïda per aquest sòl més superficial del sistema subterrani cap a zones més profundes.
- Zona saturada: és la zona d'emmagatzematge i major flux d'aigua subterrània, coneguda per la presència de formacions geològiques capacitades per albergar aigua. La línia que separa la zona no saturada de la saturada es coneix com el nivell freàtic.
- Llit rocós: zona més profunda del sistema per la qual l'aigua no circula més avall degut a la no porositat superficial de la roca.

2.1.2 Formacions geològiques subterrànies

Tot sistema d'aigües subterrànies es format per formacions geològiques que permeten l'emmagatzematge de l'aigua subterrània com a tal, originant així diferents formacions segons la capacitat d'emmagatzematge i transmissió de l'aigua, segons la porositat, segons la situació de la formació o segons la litologia.

- Segons emmagatzematge i transmissió d'aigua:
 - Aquífer: gran capacitat d'emmagatzemar i transmetre aigua.
 - Aquítard: capaç d'emmagatzemar i transmetre aigua però amb més dificultat que un aquífer.
 - Aquíclud: capaç d'emmagatzemar aigua però no transmetre-la.
 - Aquífug: incapaç d'emmagatzemar ni transmetre aigua.
- Segons la porositat:
 - Porosos.
 - Fisurats.
 - Kàrstics: Formacions geològiques carbonatades originades per meteorització química de les roques que les comprenen (calcaries, dolomites, guixos).
- Segons la situació de la formació:
 - Lliure: aquífer amb una superfície freàtica real i delimitat superiorment pel terreny i inferiorment per una capa impermeable o semipermeable.
 - Confinat: aquífer amb una superfície freàtica o superfície piezomètrica virtual que es troba limitat inferior i superiorment per capes impermeables.
 - Semiconfinat: aquífer amb superfície piezomètrica virtual limitat inferior i superiorment per una capa aquítarda la qual permet un mínim flux vertical a l'aquífer.
- Segons la litologia:
 - Detrítics.
 - Carbonatats.

2.2 Equació del flux en medi porós

La massa a un flux d'aigua és constant pel principi de conservació de la massa i suposant que l'aigua obeeix la llei de Darcy tenim com a resultat l'equació del flux.

2.2.1 Llei de Darcy

Henry Darcy va dissenyar el següent experiment de laboratori comprès per un tanc d'aigua a nivell constant connectat a un tub de secció coneguda (S) amb una sorra a l'interior on mesurava el nivell piezomètric (h_a i h_b) en dos punts A i B separats per una longitud coneguda (l) on expulsava un caudal (Q), anomenat permeàmetre.

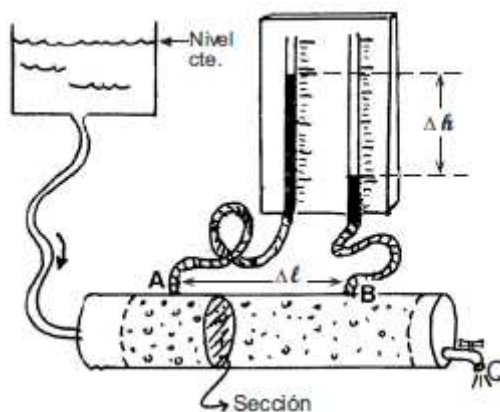


Figura 2.3. Permeàmetre utilitzat per Henry Darcy al seu experiment. (F. Javier Sánchez San Román, Juliol 2014)

Conegudes les alçades piezomètriques dels dos punts A i B i la longitud que les separa trobarem el gradient hidràulic:

$$\frac{\Delta h}{\Delta l} = \frac{h_a - h_b}{l_a - l_b} \quad (\text{Eq. 2.1})$$

Darcy va trobar que el caudal que passava pel permeàmetre era linealment proporcional a la secció i al gradient hidràulic. Si variem qualsevol variable, canviant el caudal o movent el tanc d'aigua, mesurant les alçades piezomètriques en punts diferents aquestes varien però, degut a que la sorra de dins del tub de secció coneguda no l'hem canviat sempre es compleix:

$$\frac{\Delta h}{\Delta l} S \cdot K = Q \quad (\text{Eq. 2.2})$$

On a l'Equació (Eq. 2.2):

- Δh : diferència d'alçades piezomètriques (L).
- Δl : separació longitudinal dels punts de mesura (L).
- S : Àrea de la secció (L^2).
- K : conductivitat hidràulica (L/T).
- Q : caudal (L^3/T).

Així Darcy utilitzant diferents sorres, de més fines a més gruixudes, va definir K com la permeabilitat o conductivitat hidràulica. Aquesta depèn de la connectivitat de porus i de la seva grandària.

L'expressió de la llei de Darcy final és:

$$\frac{\partial h}{\partial l} \cdot -K = q \quad (\text{Eq. 2.3})$$

On a l'Equació (Eq. 2.3) :

- q : velocitat de Darcy, Q/S caudal que circula per m^2 .
- K : conductivitat hidràulica.
- $\partial h/\partial l$: gradient hidràulic expressat en increments infinitesimals.
- El signe negatiu és degut a que el caudal és un vector amb direcció cap als Δh decreixents, el qual tindrà valor negatiu.

Parlem ara de la velocitat d'un solut en un medi porós (v) que es mou a la velocitat de l'aigua que circula, que no és la mateixa que la velocitat de Darcy (q). La relació depèn de la porositat eficaç que és la relació del volum de porus interconnectats i el volum total de la mostra.

$$\Phi_{efectiva} = \frac{\text{volum porus interconnectats}}{\text{volum total mostra}} \quad (\text{Eq. 2.4})$$

$$v = \frac{q}{\Phi_{efectiva}} \quad (\text{Eq. 2.5})$$

On:

- v és la velocitat lineal de l'aigua.
- $\Phi_{efectiva}$ és la porositat efectiva.

La llei de Darcy té unes limitacions per les quals és vàlida només si:

- Medi és homogeni.
- Medi és isòtrop.

2.2.2 Principi de conservació de la massa en un volum de control.

Suposant que és un medi isòtrop (conductivitat hidràulica constant), sabem que la massa que entra per unitat de temps i la massa que surt obtenim la variació de la massa amb el temps:

$$-\rho \left(\frac{dVx}{dx} + \frac{dVy}{dy} + \frac{dVz}{dz} \right) - \left(Vx \frac{d\rho}{dx} + Vy \frac{d\rho}{dy} + Vz \frac{d\rho}{dz} \right) = \Delta M \quad (\text{Eq. 2.6})$$

Considerem que és un medi homogeni, on la densitat ρ és constant obtenim que $\partial\rho/\partial x$, $\partial\rho/\partial y$ i $\partial\rho/\partial z$ són iguals a 0:

$$-\rho \left(\frac{dVx}{dx} + \frac{dVy}{dy} + \frac{dVz}{dz} \right) = \Delta M \quad (\text{Eq. 2.7})$$

2.2.3 Equació del flux

Un cop tenim la llei de Darcy de l'apartat 2.2.1 i el principi de conservació de la massa de l'apartat 2.2.2, si apliquem la llei de Darcy al principi de conservació de la massa:

$$\frac{dV_x}{dx} = -K \frac{d^2 h}{dx^2} \quad (\text{Eq. 2.8})$$

$$\frac{dV_y}{dy} = -K \frac{d^2 h}{dy^2} \quad (\text{Eq. 2.9})$$

$$\frac{dV_z}{dz} = -K \frac{d^2 h}{dz^2} \quad (\text{Eq. 2.10})$$

Obtenim:

$$K \nabla^2 h = (n\beta + \alpha(1 - n)) \frac{d\rho}{dt} \quad (\text{Eq. 2.11})$$

Com h és constant i la topografia no varia en el temps:

$$\frac{\partial Z}{\partial t} = 0 \quad i \quad \frac{p}{\gamma^2} \frac{\partial \gamma}{\partial t} = 0 \quad (\text{Eq. 2.12})$$

El que implica que:

$$\frac{\partial p}{\partial t} = \gamma \frac{\partial h}{\partial t} \quad (\text{Eq. 2.13})$$

Avaluant l'equació (Eq. 2.13) en l'equació (Eq. 2.11) obtenim:

$$K \nabla^2 h = (n\beta + \alpha(1 - n)) \rho g \frac{dh}{dt} \quad (\text{Eq. 2.14})$$

On:

- $S_s = (n\beta + \alpha(1 - n))$ és el coeficient d'emmagatzemat específic.

És la quantitat d'aigua emmagatzemada que s'allibera per unitat de volum quan el gradient disminueix una unitat. Els seus components s'interpreten així: $\rho g \alpha(1 - n)$ és l'aigua emmagatzemada alliberada en una unitat de volum degut a la compressió de l'esquelet intergranular quan el potencial disminueix una unitat. $\rho g n \beta$ és l'aigua emmagatzemada alliberada quan el nivell piezomètric baixa una unitat, com α és de l'ordre de $1/25 \beta$ (Bear, 1987), podem reescriure S_s com:

$$S_s = (n\beta + \alpha)\rho g \quad (\text{Eq. 2.15})$$

Pel que l'equació (Eq. 2.14) queda:

$$\nabla^2 h = \frac{S_s}{K} \frac{dh}{dt} \quad (\text{Eq. 2.16})$$

Així obtindríem l'equació general del flux, tot i que la podem expressar amb un altre paràmetre, si multipliquem a dalt i a baix per l'espessor del nostre volum de control (b) obtenim dos nous paràmetres T i S resultat de:

$$\nabla^2 h = \frac{S_s b}{K b} \frac{dh}{dt} \quad (\text{Eq. 2.17})$$

On:

$$S_s b = S \quad (\text{Eq. 2.18})$$

$$K b = T \quad (\text{Eq. 2.19})$$

- S és el coeficient d'emmagatzemat.
- T és la transmissivitat, que és el flux per unitat d'amplada a través de tot l'espessor del volum de control per a un gradient hidràulic unitari.

Resultant:

$$\nabla^2 h = \frac{S}{T} \frac{dh}{dt} \quad (\text{Eq. 2.20})$$

Capítol 3:

Introducció al transport de contaminants

3. Introducció al transport de contaminants

Quan un contaminant entra en contacte amb l'aigua subterrània es produeixen diversos processos els quals volem definir i distingir per a la posterior comprensió del comportament del contaminant.

S'ha de categoritzar depenent de dos grups:

- Contaminants reactius: el contaminant es mou segons el flux subterrani i hi ha una interacció amb el medi geològic i altres contaminants que es trobin al flux subterrani.
- Contaminants no reactius: el contaminant es mou segons el flux subterrani però no hi ha cap interacció amb el medi geològic.

3.1 Mecànica del transport en aigües subterrànies

Ara tractarem els dos grups categoritzats a l'apartat anterior, tot i que avancem que l'exclusivitat que aquest processos es produeixin de manera independent és una situació idíl·lica.

3.1.1 Processos de contaminants reactius

El transport d'un contaminant a través de les porositats d'una formació geològica dona l'oportunitat que es produeixi una interacció amb aquesta. Es poden produir processos com:

- Adsorció
- Absorció
- Intercanvi iònic
- Precipitació química
- Biodegradació
- Reaccions Redox

Aquests principals processos de contaminants reactius provoquen els següents efectes:

- Retard
- Transformació
- Atenuació

3.1.2 Processos de contaminants no reactius

Aquest processos són purament físics, deixant de banda les reaccions químiques produïdes per les interaccions amb el medi geològic.

3.1.2.1 Advecció

L'advecció és un fenomen que consisteix en l'arrossegament del contaminant per l'aigua. En un cas ideal, el contaminant viatjaria a la mateixa velocitat que l'aigua i el volum que ocupa el contaminant seria constant, on el flux de massa a través d'una secció unitat seria:

$$Flux_{advecció} = \Phi_{efectiva} v \nabla C = q \nabla C \quad (Eq. 3.1)$$

On

- $Flux_{advecció}$ és el flux de massa per unitat de secció i de temps (M/L^2T).
- $\Phi_{efectiva}$ és la porositat efectiva.
- v és la velocitat de l'aigua (M/T).
- q és la velocitat de Darcy (M/T).
- C és la concentració del contaminant (M/L^3).

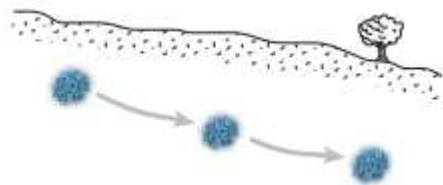


Figura 3.1. Transport únicament per advecció (F. Javier Sánchez San Román, Juny 2012).

3.1.2.2 Difusió

Si observem un volum de contaminant situat en un punt del medi i no hi hagués un flux, veuríem després d'un Δt , que el volum inicial s'ha ampliat i difuminat. Això es degut a que les molècules de contaminant dissoltes a l'aigua es mouen dels punts de major concentració a menor concentració, el qual es denominat difusió molecular o difusió.

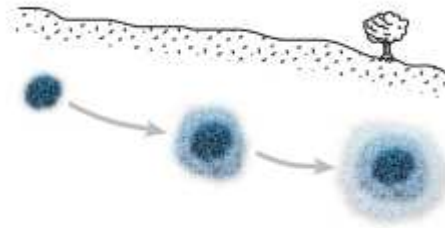


Figura 3.2. Transport combinatiu de difusió i advecció (F. Javier Sánchez San Román, Juny 2012).

En un flux de massa per difusió les molècules de contaminant es mouen per gradient de concentracions. Quan un contaminant no està contingut en un medi porós és regit per la primera llei de Fick:

$$Flux_{difusió} = -D_{difusió} \nabla C \quad (Eq. 3.2)$$

On

- $Flux_{difusió}$ és el flux de massa per unitat de secció i de temps (M/T).
- $D_{difusió}$ és el coeficient de difusió (L^2/T).
- C és la concentració del contaminant (M/L^3).

Si considerem un medi porós s'ha de considerar la porositat eficaç i la tortuositat. La tortuositat és la relació entre la longitud real que ha de recorre una partícula per anar del punt A al punt B i la distància en línia recta del punt A i el punt B. El coeficient de difusió efectiva es calcula (Fetter, 2001, 1999):

$$D'_{difusió} = D_{difusió} \cdot w \quad (\text{Eq. 3.3})$$

On

- $D'_{difusió}$ és el coeficient de difusió efectiva
- w coeficient que depèn del medi porós.

El coeficient w pot:

- Variar de 0,01 a 0,5 (Freeze i Cherry, 1979)
- Venir donat per la següent equació (Coutelieris, 2012) o (Grathwohl, 1998):

$$D'_{difusió} = \frac{D_{difusió} \cdot \Phi_{efectiva} \delta}{\tau^2} \quad (\text{Eq. 3.4})$$

on

- δ és el factor de constricció
- τ és la tortuositat
- Degut a que els coeficients δ i τ són desconeguts (Coutelieris, 2012), es simplifica la relació de la porositat efectiva:

$$D'_{difusió} = D_{difusió} \cdot \Phi_{efectiva} \quad (\text{Eq. 3.5})$$

Finalment el flux per difusió segons l'equació (Eq. 3.5), quedaria definit com:

$$Flux_{difusió} = -D'_{difusió} \nabla C \quad (\text{Eq. 3.6})$$

3.1.2.3 Dispersió mecànica

La dispersió mecànica es provocada pel moviment del fluid a través del medi porós. La dispersió es produeix en sentit longitudinal (sentit del flux) i transversalment.

- Dispersió longitudinal, suposant sempre un medi homogeni degut a que la heterogeneïtat augmenta la dispersió: es deguda a que les molècules que troben porositats més amples avancen més ràpid, les molècules que troben camins més tortuosos avancen més lentes i que les molècules que avancen pel centre dels canals intergranulars avancen més ràpid que les que circulen a prop dels grans.
- Dispersió transversal: es degut a la bifurcació constant dels camins on el fluid circula.

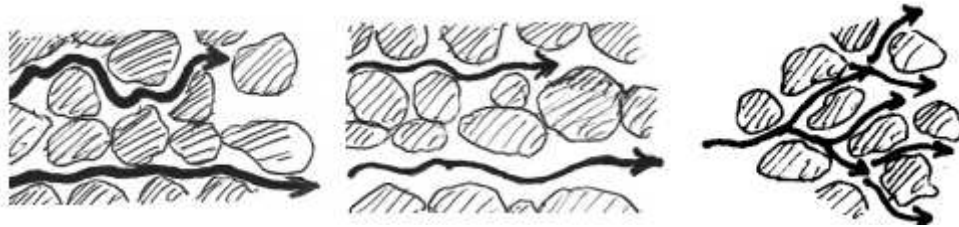


Figura 3.3. Dispersió mecànica en un medi porós. (F. Javier Sánchez San Román, Juny 2012).

La capacitat del medi porós per dispersar mecànicament es coneix com dispersió dinàmica (α), diferenciant així la dispersió dinàmica longitudinal (α_L) i la dispersió dinàmica transversal (α_T).

La dispersió mecànica és igual a:

$$D_{mecànica} = \alpha \cdot v \quad (\text{Eq. 3.7})$$

On

- α és la dispersió dinàmica (L)
- v és la velocitat lineal mitja (L/T)

El flux de la dispersió dinàmica és:

$$Flux_{dispersió\ mecànica} = -D_{mecànica} \Phi_{efectiva} \nabla C = -D'_{mecànica} \nabla C \quad (\text{Eq. 3.8})$$

On

- Flux_{dispersió mecànica} és el flux de massa per dispersió mecànica per unitat d'àrea i temps (M/L²T).
- D' mecànica és el coeficient de dispersió mecànica.

3.1.2.4 Dispersió hidrodinàmica

És l'acció conjunta de la difusió i la dispersió mecànica on la suma dels coeficients de difusió i dispersió mecànica dóna el coeficient de dispersió hidrodinàmica ($D_{d.hidrodinàmica}$). Tot i que segons la direcció del flux (longitudinal o transversal) podem expressar-ho segons el mòdul de la velocitat:

$$D_{disp\ L} = D_{difusió} + \alpha_L \cdot v \quad (\text{Eq. 3.9})$$

$$D_{disp\ T} = D_{difusió} + \alpha_T \cdot v \quad (\text{Eq. 3.10})$$

On

- $D_{disp\ L}$ és el coeficient de dispersió longitudinal (L²/T).
- $D_{disp\ T}$ és el coeficient de dispersió transversal (L²/T).
- α_L és el coeficient de dispersió mecànica longitudinal (L).
- α_T és el coeficient de dispersió mecànica transversal (L).

Es defineix el flux de dispersió hidrodinàmica com:

$$\begin{aligned} Flux_{dispersió\ hidrodinàmica} &= Flux_{dispersió\ mecànica} + Flux_{difusió} = \\ &= D_{d.hidro} \Phi_{efectiva} \nabla C \end{aligned} \quad (\text{Eq. 3.11})$$

On

- *Flux dispersió hidrodinàmica* és el flux màssic per dispersió. hidrodinàmica (M/L²T).

3.1.3 Equació Advecció- Dispersió

La suma dels fluxos dels processos anteriors obtenim el flux màssic (Eq. 3.12), on si substituïm les equacions (Eq. 3.1), (Eq. 3.2), (Eq. 3.6), (Eq. 3.8), obtenim l'equació d'advecció-dispersió (ADE) (Eq. 3.13).

$$Flux_{màssic} = Flux_{advecció} + Flux_{difusió} + Flux_{dispersió\ mecànica} \quad (Eq. 3.12)$$

$$-\nabla[(q\nabla C) + (-D'_{difusió}\nabla C) + (-D'_{mecànica}\nabla C)] = \frac{\partial C}{\partial t} \quad (Eq. 3.13)$$

3.2 Mètodes numèrics aplicats al transport de contaminants

La selecció d'una aproximació adequada per resoldre l'equació del transport de contaminants és de vital importància, degut a la gran varietat de mètodes, on les característiques de cadascun donaran la precisió i eficiència de l'aproximació del model a la realitat.

3.2.1 Mètodes Eulerians

Els mètodes Eulerians són els mètodes més antics i comuns per la resolució de l'equació del transport de contaminants. Dóna aproximacions de les discretitzacions de l'ETC en una malla fixa. La contra d'aquests mètodes és la dispersió numèrica que produeix en problemes de gran advecció. Les formulacions més utilitzades són el mètode de diferències finites (MDF), el mètode d'elements finits (MEF) i el mètode de volums finits (MVF).

3.2.2 Mètodes Lagrangians

A diferència dels mètodes Eulerians, els mètodes Lagrangians no utilitzen una malla fixa per discretitzar l'ETC, sinó que utilitzen una malla deformable sobre un sistema de coordenades fix. Aquests mètodes donen bons resultats tot i que tenen un alt cost computacional. Els mètodes lagrangians més utilitzats són *Random Walk Particle Tracking* (RWPT) i *Continuous Time Particle Walk* (CTPW).

3.2.3 Mètodes Mixtes (Eulerià-Lagrangià)

Els mètodes mixtes resolen l'ETC sobre una malla fixa, on es divideix el procés en dues fases: a la primera fase les concentracions als nodes es desplacen a un node posterior seguint la direcció del flux, i a la segona fase, es calcula l'efecte de la dispersió. Alguns mètodes mixtes són Mètode de les Característiques (MOC) o el Mètode de les Característiques Modificades (MMOC).

Capítol 4:

Particle Tracking Method (PTMs)

4. Particle Tracking Methods (PTMs)

Els PTMs són mètodes Lagrangians, molt versàtils en moltes aplicacions. Són òptims pel càlcul de problemes dominats per l'advecció en medis heterogenis. L'inconvenient és que els PTMs converteixen les distribucions de partícules en concentracions, on ens trobem amb el repte de l'inferència d'una concentració localitzada a les *breakthrough curves* (BTCs) de diferents temps provinents de les partícules.

Les concentracions reals només les podem obtenir en el límit, quan un infinit número de partícules han estat utilitzades, per aquest motiu sempre estimarem un error a la solució final.

La definició de les BTCs és molt important en el transport de contaminants, on l'arribada de contaminants ens dona l'idea del potencial d'aquest, però l'arribada de les últimes concentracions són claus per determinar quina és l'estada del contaminant al teu aquífer. Els PTM's en un cas real, ofereixen BTCs asimètriques a llarga durada, les quals veiem una cua irregular, la que no dona un model de concentracions òptim a llarg termini, el qual s'anomena *tailing*. Aquest comportament és esmenat com *heavy-tailing BTCs*. L'òptima reconstrucció de les *heavy-tailing BTCs* és corregit pels PTM's amb un gran número de partícules per descriure la distribució de concentració sencera.

4.1 Random Walk Particle Tracking (RWPT)

El *Random Walk Particle Tracking* (RWPT) és un PTM que analitza el transport de contaminants mitjançant conceptes estadístics, el qual representa les concentracions a través de les partícules on cada partícula posseeix una part de la massa total del contaminant.

L'equació del transport de contaminants conservatius a un medi heterogeni ve donada per:

$$\frac{\partial C}{\partial t} + \nabla(qC) = \nabla(D\nabla C) \quad (\text{Eq. 4.1})$$

On

- D és el tensor de dispersió mecànica.

L'evolució de la partícula en el temps ve donada pels termes de moviment advection i moviment Brownian responsable de la dispersió. El desplaçament ve donat per l'esquema d'integració de Itô-Taylor (Gardiner, 1990):

$$X_p(t + \Delta t) = X_p(t) + A(X_p, t)\Delta t + B(X_p, t) \cdot \varepsilon(t)\sqrt{\Delta t} \quad (\text{Eq. 4.2})$$

On

- $X_p(t)$ és la posició de la partícula en el temps t .
- A és el vector de deriva
- B és la matriu de desplaçament
- $\xi(t)$ és un vector independent de variables aleatòries de mitja zero i variància u.

La funció de distribució de densitat $f(X_p(t))$ és la distribució de la probabilitat de trobar una partícula dins d'un interval donat $(X_p, X_p + dX_p)$ en un temps t , obtingut de (Eq. 4.2) (Itô, 1951), compleix amb l'equació de *Fokker-Planck* (Eq. 4.3) al límit d'un gran nombre de partícules i un Δt infinitesimal. (Eq. 4.3) descriu el moviment de la distribució de densitat (f) de les partícules:

$$\frac{\partial f}{\partial t} + \nabla(qf) = \nabla \nabla : (Df) \quad (\text{Eq. 4.3})$$

L'equació de *Fokker-Planck* (Eq. 4.3) i l'equació d'advecció-dispersió (Eq. 3.13) són molt similars, ambdues compostes de termes d'advecció, dispersió i difusió. Per establir una relació entre elles, l'equació del transport de contaminants s'ha de modificar (Kinzelbach. W, 1987):

$$\frac{\partial C}{\partial t} + (q \nabla C) + (C \nabla D) = \nabla \nabla : (D_c) \quad (\text{Eq. 4.4})$$

Modificant la velocitat,

$$q' = q + \nabla D \quad (\text{Eq. 4.5})$$

Es demostra que l'equació de transport de contaminants (Eq. 4.1), es pot transformar en una equació equivalent a l'equació de Fokker-Planck (Eq. 4.3):

$$\frac{\partial C}{\partial t} + \nabla(qC) = \nabla \nabla : (DC) \quad (\text{Eq. 4.6})$$

Si substituïm el vector de velocitat modificada (Eq. 4.5), en el valor A , de (Eq. 4.2), l'esquema de trajectòries és:

$$X_p(t + \Delta t) = X_p(t) + (q(X_p, t) + \nabla D(X_p, t))\Delta t + B(X_p, t) \cdot \varepsilon(t)\sqrt{\Delta t} \quad (\text{Eq. 4.7})$$

On la matriu de desplaçament B està relacionada amb el tensor de dispersió D :

$$2D = B \cdot B^T \quad (\text{Eq. 4.8})$$

On D està definida amb q i no q' . La forma tridimensional de la matriu de desplaçament B per a medis porosos e isotròpics (Salamon, P. Et al, 2006) és:

$$B = \begin{pmatrix} \frac{v_x}{|v|} \sqrt{2(\alpha_L|v| + D_m)} & -\frac{v_x \cdot v_z}{|v| \sqrt{v_x^2 + v_y^2}} \sqrt{2(\alpha_T|v| + D_m)} & -\frac{v_y}{\sqrt{v_x^2 + v_y^2}} \sqrt{2(\alpha_T|v| + D_m)} \\ \frac{v_y}{|v|} \sqrt{2(\alpha_L|v| + D_m)} & -\frac{v_y \cdot v_z}{|v| \sqrt{v_x^2 + v_y^2}} \sqrt{2(\alpha_T|v| + D_m)} & -\frac{v_x}{\sqrt{v_x^2 + v_y^2}} \sqrt{2(\alpha_T|v| + D_m)} \\ \frac{v_z}{|v|} \sqrt{2(\alpha_L|v| + D_m)} & -\frac{\sqrt{v_x^2 + v_y^2}}{|v|} \sqrt{2(\alpha_T|v| + D_m)} & 0 \end{pmatrix} \quad (\text{Eq. 4.9})$$

4.2 Kernel Density Estimators (KDE)

La reconstrucció de les BTCs és la major desavantatge dels PTMs. En el procés de reconstrucció d'una BTC, on es dedueixen les concentracions són proporcionals a les probabilitat de densitat de partícules (PDFs), les concentracions de flux estan relacionades amb la PDF en el temps amb:

$$f(t) = \frac{Q}{m_0} c_f \quad (\text{Eq. 4.10})$$

$$m_0 = \int_0^{\infty} Q c_f dt \quad (\text{Eq. 4.11})$$

On

- $f(t)$ són les densitats probables de partícules en el temps.
- Q és el flux total.
- c_f són les concentracions de flux.
- m_0 és la massa de partícules situada per sota de la BTC.

Aquesta reconstrucció és difícil, ja que mai obtindrem un nombre infinit de partícules per poder validar les equacions (Eq. 4.10), (Eq. 4.11).

Els Kernel Density Estimators (KDE) donen una opció vàlida per la reconstrucció de les PDFs. Ofereixen una millora de l'estimació de $f(t)$, on l'estimació de les PDFs és determinava amb histogrames de freqüència, on es compta el número de partícules en un interval de temps $B_j = [t_0 + (j-1)h, t_0 + jh]$, on $f(t)$ en el moment j ve determinat per:

$$f_j \approx \hat{f}_j = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} I\{t_i \in B_j\} \quad (\text{Eq. 4.12})$$

On

- n és el número de partícules.
- \hat{f}_j és la densitat estimada.
- $I\{\cdot\}$ és un indicador, el qual $I=1$ si t_i és dins de B_j , si no, $I=0$.
- h és l'interval de temps.

L'histograma és un model de caixes, que depèn de t_0 , de h , les quals són aleatòries. Els KDE solucionen aquest problema, canviant el model de caixes de l'histograma per una funció ponderada K centrada a la localització de la partícula:

$$f(t) \approx \hat{f}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - t_i}{h}\right) \quad (\text{Eq. 4.13})$$

On

- h és el coeficient de *bandwidth*.
- K és la funció ponderada.

Les funcions típiques de *kernel*, venen recollides a *Taula 4.1*, on el *kernel* Gaussià és més utilitzat degut a les avantatges matemàtiques.

Taula 4.1 Exemples de la funció del Kernel (Silverman, 1986)

TIPUS DE KERNEL	$K(\tau)$
Biweight	$\begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{\tau^2}{5}\right) & \text{for } \tau < 1 \\ 0 & \text{otherwise} \end{cases}$
Epanechnikow	$\begin{cases} \frac{15}{16} (1 - \tau^2)^2 & \text{for } \tau < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$
Gaussian	$(2\pi)^{-1/2} \exp\left(\frac{-\tau^2}{2}\right)$
Rectangular	$\begin{cases} \frac{1}{2} & \text{for } \tau < 1 \\ 0 & \text{otherwise} \end{cases}$
Triangular	$\begin{cases} 1 - \tau & \text{for } \tau < 1 \\ 0 & \text{otherwise} \end{cases}$

El comportament de $\hat{f}(t)$ és analitzat assumint que les dades són independents a la funció de probabilitat de densitats $f(t)$ i veient que $\hat{f}(t)$ és un estimador de $f(t)$. Una manera d'analitzar l'error d'aquesta estimació és amb el *mean integrated squared error* (MISE):

$$MISE(h) = E \int (\hat{f}_h - f)^2 \quad (\text{Eq. 4.14})$$

On podem afegir que:

$$MISE(h) = E \int \text{bias}^2(\hat{f}_h) dt + \int \text{var}(\hat{f}_h) dt \quad (\text{Eq. 4.15})$$

On

- $\text{Bias}(\hat{f}) = E(\hat{f}) - f$
- $\text{var}(\hat{f}) = E([\hat{f} - E(\hat{f})]^2)$

L'anàlisi quantifica com funciona el coeficient h com un paràmetre suavitzador. En suposicions estàndard, el MISE és asimptòtic ($n \rightarrow \infty$), aproximat pel *Asymptotic MISE* (AMISE) (Silverman, 1986) :

$$AMISE(h) = n^{-1}h^{-1}R(K) + h^4R(f'') \left(\int x^2 \frac{K}{2} \right)^2 \quad (\text{Eq. 4.16})$$

En (Eq. 4.17) el primer terme (variància) és gran quan h és molt petita, i el segon terme (*bias*) és gran quan h és molt gran. La quantificació del efecte de suavitzat de h , si minimitzem (Eq. 4.16), expresseu h com:

$$h_{AMISE} = \left[\frac{R(K)}{nR(f'') \left(\int x^2 K \right)^2} \right]^{\frac{1}{5}} \quad (\text{Eq. 4.17})$$

4.3 Bandwidth (h)

L'elecció d'un coeficient de bandwidth adient pot esdevenir un problema en els KDE. Quan un bandwidth és molt baix, el càlcul de densitat de partícules és abrupte i pot esdevenir densitats poc acurades o falses. Pel contrari, quan el bandwidth és molt alt, l'estructura de la BTC a llarg termini es pot veure suavitzada, perdent informació real de les densitats.

L'elecció manual d'un BW és poc viable, degut a la necessitat de l'ajust del BW a les teves dades a tractar. Si s'obté una BTC que es podrà tractar i discutir, però les dades seran irreals degut a la falta d'experiència amb aquestes. Un mètode automàtic estalviarà temps i ens donarà precisió.

A continuació detallarem tres mètodes pel càlcul del BW.

4.3.1 Rules of Thumb

És el mètode més senzill per l'elecció d'un BW global. S'identifica per la substitució de les parts desconegudes de (Eq. 4.17), $R(f'')$ per una desviació estàndard normal $N(0, \sigma^2)$. Els BW calculats amb aquesta tècnica són coneguts normalment per ser més gran que el BW ideal, el qual produiria una suavització de la BTC calculada. De les diverses interpretacions d'aquest mètode, ens quedem amb la de *Silverman's Rule of Thumb* (Silverman, 1986):

$$h_{SNR} = 1.06 S n^{-\frac{1}{5}} \quad (\text{Eq. 4.18})$$

On

- S és la desviació estàndard.

Posteriorment, Silverman (1986), va recomanar el canvi de la desviació estàndard per el rang interquartil dividit per 1.34, on l'expressió del BW va quedar:

$$h_{ROT} = 1.06An^{\frac{-1}{5}} \quad (\text{Eq. 4.19})$$

On

- A és el rang interquartil dividit per 1.34.

Altres maneres de calcular aquest BW, és l'expressada per Terrel and Scott (1985) i Terrel (1990), on van desenvolupar un BW amb el màxim paràmetre de suavitzat, que s'aconseguia reduït el paràmetre $R(f'')$, al mínim valor:

$$h_{OS} = 1.144Sn^{\frac{-1}{5}} \quad (\text{Eq. 4.20})$$

4.3.2 Least Squares Cross-Validation

L'idea proposada per Rudemo(1982) i Bowman(1984), contempla estimar el BW amb la minimització de l'*integrated square error* (ISE), on el criteri a seguir per la minimització és:

$$ISE(h) = \int (\hat{f}_h - f)^2 \quad (\text{Eq. 4.21})$$

$$h_{LSCV} = \underset{h>0}{\operatorname{argmin}} LSCV(h, r) \quad (\text{Eq. 4.22})$$

On $LSCV(h, r)$ és:

$$LSCV(h, r) = \frac{R(K^r)}{nh^{2r+1}} + \frac{(-1)^r}{n(n-1)h^{2r+1}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n (K^r * K^r - 2K^{2r}) \left(\frac{X_j - X_i}{h} \right) \quad (\text{Eq. 4.23})$$

El BW calculat per aquest mètode tendeix a donar valors baixos, on les BTCs no es mostren gaire suavitzades on trobaríem densitats errònies allunyades de la real.

4.3.3 Solve-the-Equation Plug-in Approach

L'idea principal és introduir un estimador de $R(f'')$, en h_{AMISE} . Aquesta aproximació proposada per Hall(1980) i Sheather (1983, 1986), ha estat acurada posteriorment per Sheather and Jones(1991):

$$h_{SPI} - \left[\frac{1}{\sqrt{2\pi}\hat{\Phi}_4[\gamma(h)]N} \right]^{\frac{1}{5}} = 0 \quad (\text{Eq. 4.24})$$

On $\hat{\Phi}_4[\gamma(h)]$ és l'estimador de la variància al quadrat, definit en (Eq. 4.25):

$$\hat{\Phi}_4[\gamma(h)] = \frac{1}{N(N-1)\sqrt{2\pi}\gamma(h)^5} \sum_{i=1}^N \sum_{j=1}^N H_4\left(\frac{x_i - x_j}{\gamma(h)}\right) e^{-(x_i - x_j)/2\gamma(h)^2} \quad (\text{Eq. 4.25})$$

On $\gamma(h)$ és:

$$\gamma(h) = \left[\frac{-6\sqrt{2}\hat{\Phi}_4(g_1)}{\hat{\Phi}_6(g_2)} \right]^{\frac{1}{5}} h^{\frac{5}{7}} \quad (\text{Eq. 4.26})$$

Aquesta equació pot ser resolta per un mètode numèric com el de *Newton-Rapshon*.

Capítol 5:

MATLAB: Resolució de densitats de partícules

5. MATLAB: Resolució de densitats de partícules

A continuació veurem el codi desenvolupat per la resolució de densitat de partícules a partir del model de KDE en una dimensió, on el BW és un valor escalar, obtingut mitjançant la *Rule of Thumb de Silverman (1986)* i *Solve the Equation Plug-in approach*, i el vector de dades és un vector aleatori, d'una distribució coneguda. El codi crida una funció anomenada *kde.m*, la qual fa un tractament del vector de dades per obtenir un rang de les dades màxim, i després calcula la funció de densitats estimades. Crida una altra funció anomenada *BW.m* que calcula el bandwidth pel mètode *Solve the Equation Plug-in approach*.

```
close all
clear all

% Introduïm els INPUTS per aquest Model: %
% Creem una mostra aleatòria de caire (1), (2), (3), (4) on: %
% (1) Distribució Normal %
% (2) Distribució Lognormal %
% (3) Distribució Beta %
% (4) Distribució Bimodal %
% Amb els diferents paràmetres de les distribucions pertinents son: %
% a) Tingui un nombre "N" de partícules %
% b) On "u" sigui el valor de la mitja %
% c) "var" satisfaci el valor de la variància %
% d) A i B paràmetres de la distribució Beta %
%
% Calculem el BW segons quatre mètodes: %
% Rule of Thumb:
%     Silverman's Normal Reference %
%     Terrel and Scoot Over Smoothed %
%     Silverman's Rule of Thumb %
% Solve the Equation Plug in Method %
% Donem un nombre de punts on avaluar la funció de densitat %
%
% Paràmetres %
N1 = 100;
u1 = 1;
var1 = 0.1;
N2 = 600;
u2 = 5;
var2 = 0.5;
% A=1;
% B=5;
%
% x1 = u1 + sqrt(var1)*randn(N1,1); % Distribució normal
% x1 = lognrnd(u1,var1,N1,1); % Distribució lognormal
% x1 = betarnd(A,B,N1,1); % Distribució beta
x = u1 + sqrt(var1)*randn(N1,1);
x2 = u2 + sqrt(var2)*randn(N2,1); % Distribució bimodal
x1 = [x;x2];
%
```

```

h1=1.06*(iqr(x1)/1.34)*(N1)^(-1/5);%Silverman's Rule of thumb %
h2=1.144*std(x1)*(N1)^(-1/5); % Terrel and Scott Over smoothed %
h3=1.06*std(x1)*(N1)^(-1/5); % Silverman's normal reference %
%
j=4; % Solve the equation plug-in method %
h4=BW(x1,j); % Cridem a la funció BW que executa el càlcul %
%
npunts = 10;
% npunts2=10000;
% npunts3=100000;
%
%Cridem la funció kde, la qual ens calcularà els valors de les
densitats
%estimates pèl mètode %
[dades,densitats] = kde(x1,h1,npunts,1.2);
[dades2,densitats2] = kde(x1,h2,npunts,1.2);
[dades3,densitats3] = kde(x1,h3,npunts,1.2);
[dades4,densitats4] = kde(x1,h4,npunts,1.2);
%
% Obtenim els resultats: %
% Representem la funció i la comparem amb l'histograma que obtindriem
de
% les mateixes dades %
figure(1);
plot(dades,densitats,'b','LineWidth',2);
hold on
plot(dades2,densitats2,'r','LineWidth',2);
hold on
plot(dades3,densitats3,'k','LineWidth',2);
hold on
plot(dades4,densitats4,'y','LineWidth',2);
hold on
[n,q] = hist(x1,N1);
bar(q,n/sum(n)/abs(q(2)-q(1)),'k','g')
title('Densitat estimada de partícules')
xlabel('distància')
ylabel('Concentracions')
legend('Hrot','Hos','Hsnr','Plug-in','Densitats partícules')
hold off

figure(2)
plot(dades,densitats,'b','LineWidth',2);
hold on
[n,q] = hist(x1,N1);
bar(q,n/sum(n)/abs(q(2)-q(1)),'k','g')
hold off
title('Densitat estimada de partícules')
xlabel('distància')
ylabel('Concentracions')
legend('Hrot')

figure(3)
plot(dades2,densitats2,'r','LineWidth',2);
hold on
[n,q] = hist(x1,N1);
bar(q,n/sum(n)/abs(q(2)-q(1)),'k','g')
title('Densitat estimada de partícules')
xlabel('distància')
ylabel('Concentracions')
legend('Hos')
hold off

```

```
figure(4)
plot(dades3,densitats3,'k','LineWidth',2);
hold on
[n,q] = hist(x1,N1);
bar(q,n/sum(n)/abs(q(2)-q(1)),'k','g')
hold off
title('Densitat estimada de partícules')
xlabel('distància')
ylabel('Concentracions')
legend('Hsnr')
```

```
figure(5)
plot(dades4,densitats4,'y','LineWidth',2);
hold on
[n,q] = hist(x1,N1);
bar(q,n/sum(n)/abs(q(2)-q(1)),'k','g')
hold off
title('Densitat estimada de partícules')
xlabel('distància')
ylabel('Concentracions')
legend('Hste')
```

La funció *kde.m* és:

```
function [x1,y] = kde(valors,h,npunts,rangx)

% Aquesta funció ens calcula les densitats estimades del vector de
dades
% aleatòries "x" %
%
% Busquem el número de valors (dades) que tenim %
N = length(valors);
%
%Allarguem el vector "valors" per abastir tot el mallat on utilitzem
un factor 1.2%
mm = minmax(valors(:));
if length(npunts) > 1
    x1 = npunts;
else
    xm = rangx*mm;
    if (mm(1)>0),
        xm(1)=mm(1)/rangx;
    end

    if (mm(2)<0),
        xm(2)=mm(2)/rangx;
    end
    x1 = linspace(xm(1),xm(2),npunts);
end

%
% Deixem córrer la funció que ens donarà les dades de densitats
respecte la mostra aleatòria de partícules %
%
% "y" es la funció de densitats estimades %
%
y = zeros(size(x1));
```

```
c = 1/(sqrt(2*pi)*h);
for i=1:N,
    y = y + 1/N*c*exp(-(valors(i)-x1).^2/(2*h));
end
```

La funció *BW.m* és:

```
function h1=BW(x1,j)
% Càlcul del BW mitjançant Solve the Equation Plug-in method %
n=length(x1);
s=x1*ones(1,length(x1))-ones(length(x1),1)*x1';
r=gamma(j+1.5)/2/pi/(std(x1).^(2*j+3));
for m=j:-1:1
    a=(gamma(m+.5)*(2^(m+.5))/pi/r/n)^(1/(2*m+3));
    u=s/a;
    h0=1;
    h1=u;
    for i=1:2*m-1
        hr=u.*h1-i*h0;
        h0=h1; h1=hr;
    end;
    r=abs(mean(mean(normpdf(u,0,1).*hr)'))/(a^(1+2*m));
end;
h1=1/((sqrt(pi)*r*n)^(1/3));
```

Capítol 6:

Resultats

6. Resultats

Ara mostrarem els resultats obtinguts. Al codi de MATLAB hi ha quatre exemples exposats, on varia la distribució que segueixen les partícules de contaminant. Les mostrades són:

- Distribució Normal
- Distribució Lognormal
- Distribució Beta
- Distribució Bimodal

Així mateix, tractarem les diferents distribucions segons els resultats obtinguts i segons dos paràmetres, un, el número de punts en els que analitzem la funció, on la quantitat de partícules utilitzades al mètode ens determinarà la precisió del model. El número de punts on les analitzem s'esdevindrà de caràcter més o menys òptim. Parlem del mallat característic que presenta aquest tipus de mètodes Lagrangians i el número de punts que disposin. L'altre paràmetre és el BW escollit, com s'ha esmentat al apartat 4.3, és un paràmetre suavitzador de la funció de densitats estimades.

Per últim, analitzarem l'error de cada mètode comparant-ho amb la solució analítica de cada distribució.

6.1 Distribució Normal

La distribució normal o de Gauss, de variable continua aleatòria és de les més comuns que ens trobem. La seva principal característica és la simetria que aquesta presenta, factor pel qual l'hem escollit pel càlcul de densitats estimades.

Ara tractarem unes dades aleatòries de distribució normal, i les discutirem segons el número de punts i el BW.

6.1.1 Distribució Normal: Número de punts

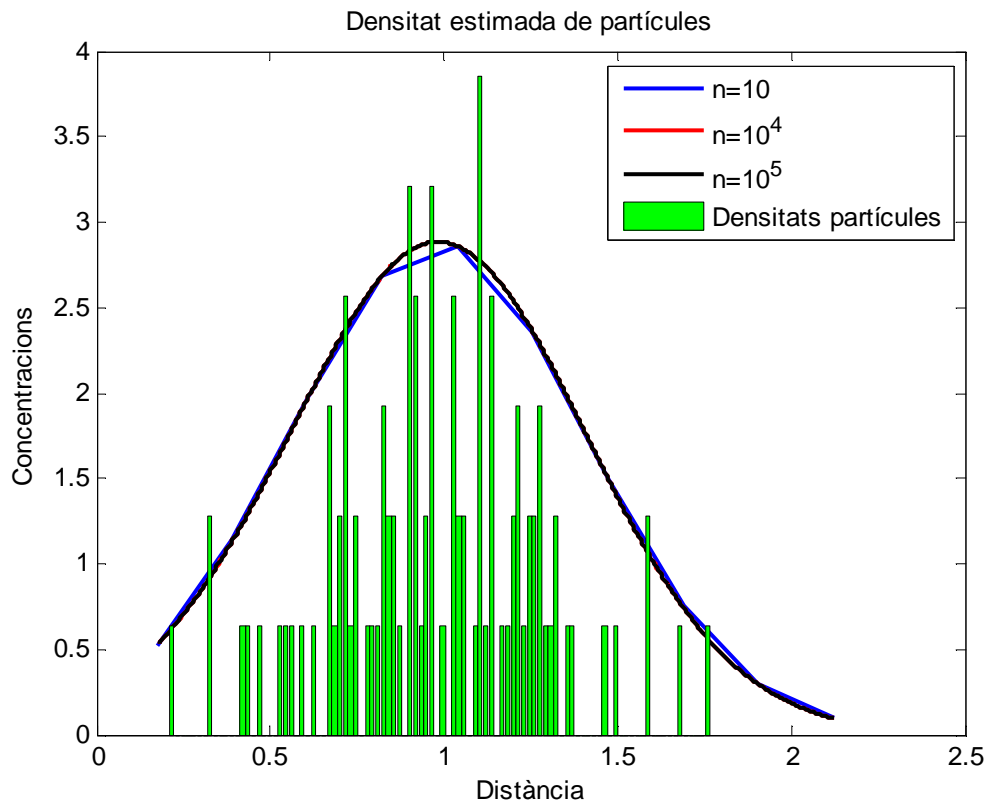


Figura 6.1. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució normal. (Extreta de MATLAB)

A la *Figura 6.1*, observem com per una mateixa distribució, amb la variació del número de punts del nostre mallat característic obtenim una aproximació mes o menys òptima de la funció de densitats estimades de les partícules. Com es veu a la funció de color blau ($n=10$), la manca de punts és visible degut a la rigidesa de la gràfica, efecte que no observem a les altres dues gràfiques. Per la funció vermella ($n=10^4$) i la funció negra ($n=10^5$), la diferència és mínima degut a la poca variància o extensió de les dades del mostreig ($\text{var}=0.1$).

Comparem també les distribucions amb l'histograma que obtindríem de les dades, on observem la millora del mètode KDE respecte els histogrames. Veiem una continuïtat de les densitats en l'espai, més enllà de la unió d'una columna de l'histograma amb la següent, sinó rebent un tractament continu de les dades.

6.1.2 Distribució Normal: Bandwidth

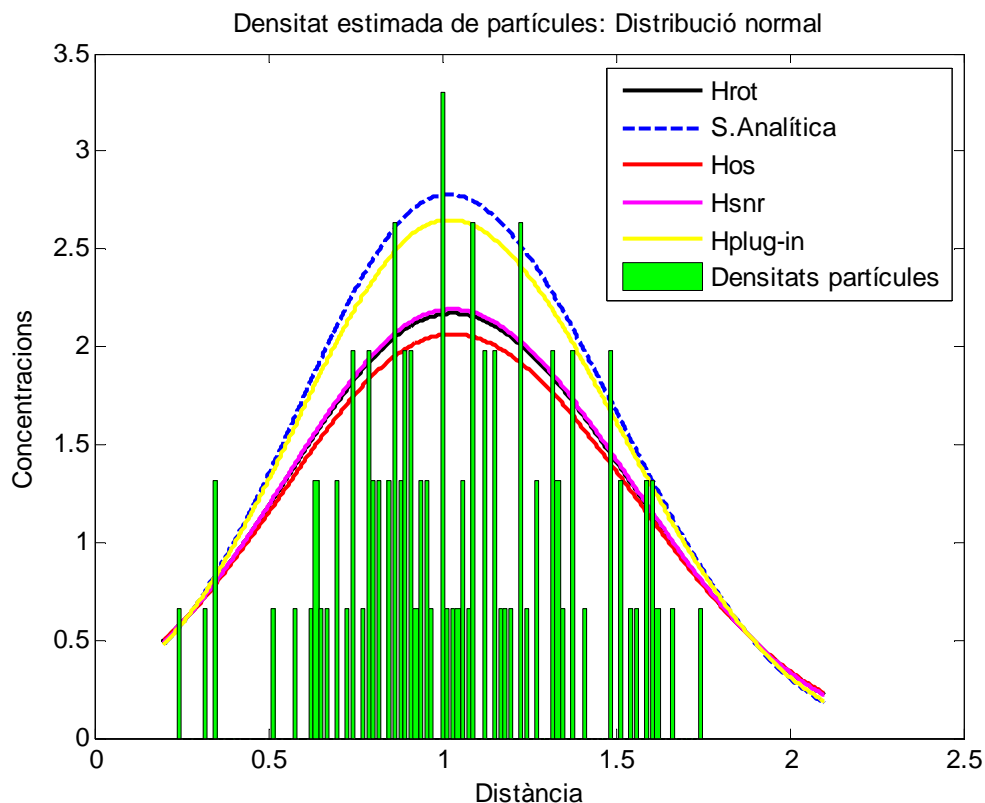


Figura 6.2. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució normal. (Extreta de MATLAB).

Com s'ha esmenat a l'apartat Bandwidth (h), els mètodes per calcular el BW són diversos. Podem fer una selecció entre sobre-suavitats i infra-suavitats, els quals indiquen l'apropament a la funció de densitats real. H_{os} , H_{rot} i H_{snr} són valors de BW sobre suavitzats, els quals són bons valors per una idea inicial, degut al baix cost computacional, però no més enllà d'aquesta. $H_{plug-in}$ és un mètode més sofisticat el qual el paràmetre és més fidel a la realitat, el que es pot prendre seriosament i rebre un tractament de les dades òptim pel desenvolupament d'un càlcul. En aquesta distribució normal, s'observa el gradient de funcions de densitats des del paràmetres més sobre-suavitat (H_{os}) fins al més òptim ($H_{plug-in}$).

6.1.3 Distribució Normal: Error

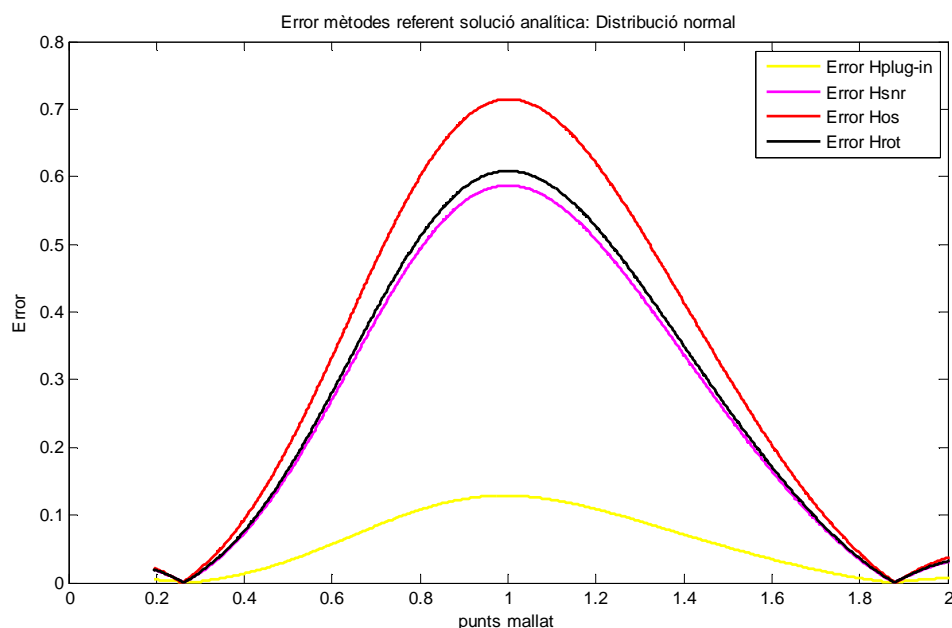


Figura 6.3. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució normal, (Extreta de MATLAB).

Observem doncs, que en aquest primer cas amb una distribució simètrica, l'error del mètode amb el BW més òptim ($H_{\text{plug-in}}$), és aproximat a la dècima en el seu valor més alt. En la resta de casos hi ha un increment de l'error significatiu parlant dels valors normals de contaminants en aigües subterrànies.

6.2 Distribució Lognormal

La distribució és caracteritza com una distribució normal, amb la principal diferència de la asimetria de les dades degut a la multiplicació per una constant de les dades normalment distribuïdes.

Els paràmetres de mitja i variància escollits són 1 i 0.35, respectivament. On emulem una distribució asimètrica, extreta de dades simètriques pertanyents a una distribució normal.

6.2.1 Distribució Lognormal: Número de punts

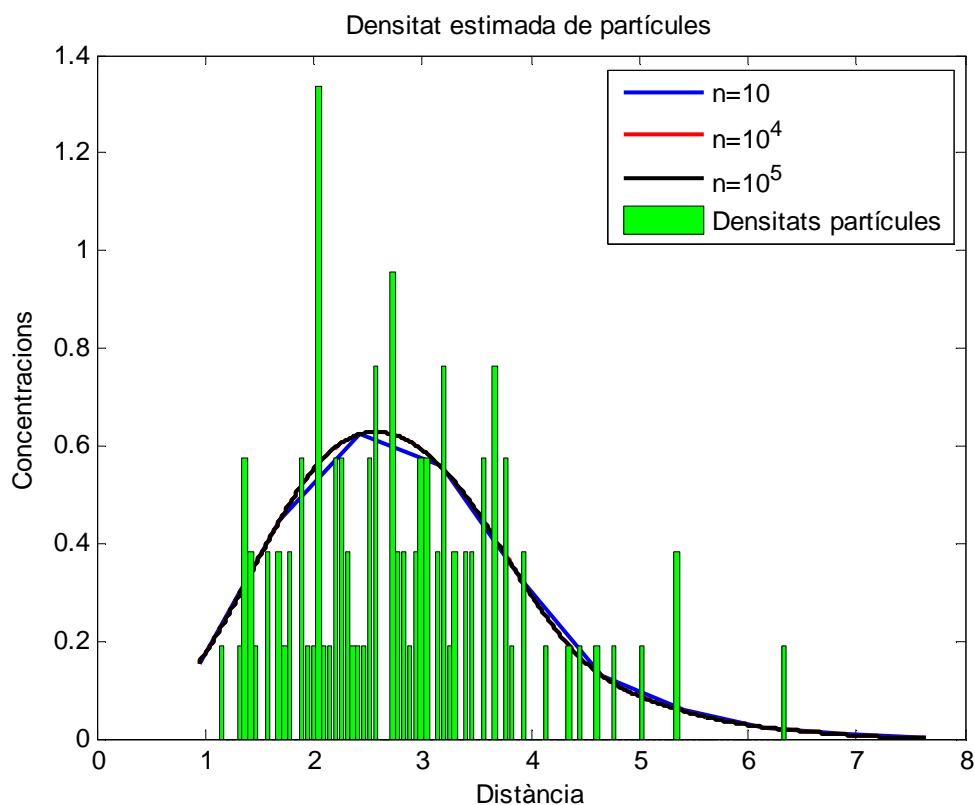


Figura 6.4. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució lognormal. (Extreta de MATLAB).

Al igual que en la distribució normal, s'observa la manca de continuïtat del primer cas ($n=10$), i la gran similitud entre els altres dos casos ($n=10^4$) i ($n=10^5$).

La diferència en aquestes dues distribucions com s'ha esmentat és la asimetria que presenta aquesta, apropant-nos més a distribucions de contaminants reals que en el cas simètric.

6.2.2 Distribució lognormal: Bandwidth

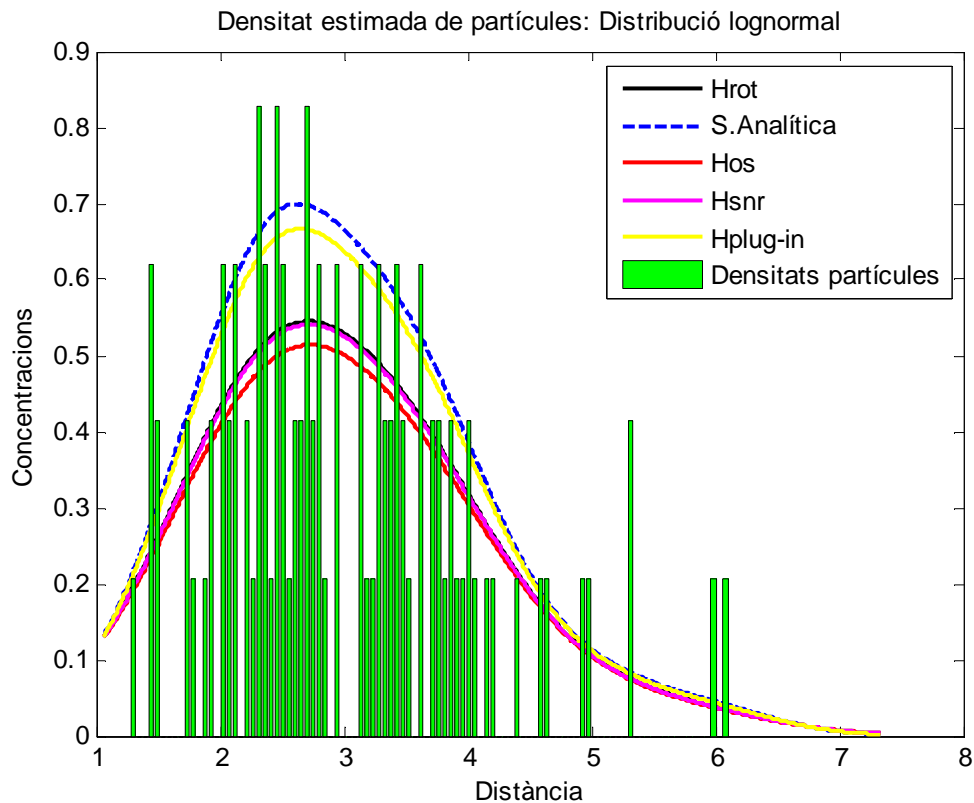


Figura 6.5. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució lognormal. (Extreta de MATLAB).

En aquest cas, la distribució és asimètrica encara que prové de d'una distribució simètrica com ho és a la distribució normal, observem la similitud amb el cas anterior. Un paràmetre ($H_{\text{plug-in}}$) com a més òptim seguit de la resta.

6.2.3 Distribució Lognormal: Error

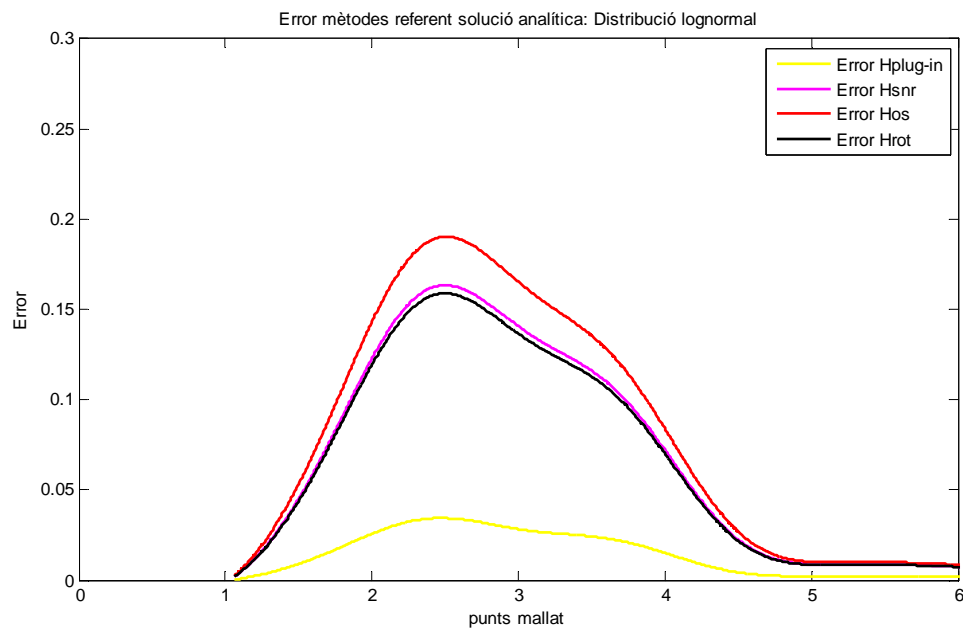


Figura 6.6.. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució lognormal, (Extreta de MATLAB).

L'error en aquesta distribució lognormal és similar al cas normal, on el màxim el trobem als bandwidth més sobre suavitzats i els més òptims als bandwidth més elaborats com en el cas del *plug-in method*.

6.3 Distribució Beta

La distribució Beta és una funció de variable continua caracteritzada per dos paràmetres A i B.

Els paràmetres A i B escollits són 1 i 3, respectivament. Així emulem un transport de contaminants amb un pic de densitat màxim molt proper i una cua ben definida on podem veure la degradació d'aquest en el nostre mallat de control.

6.3.1 Distribució Beta: Número de punts

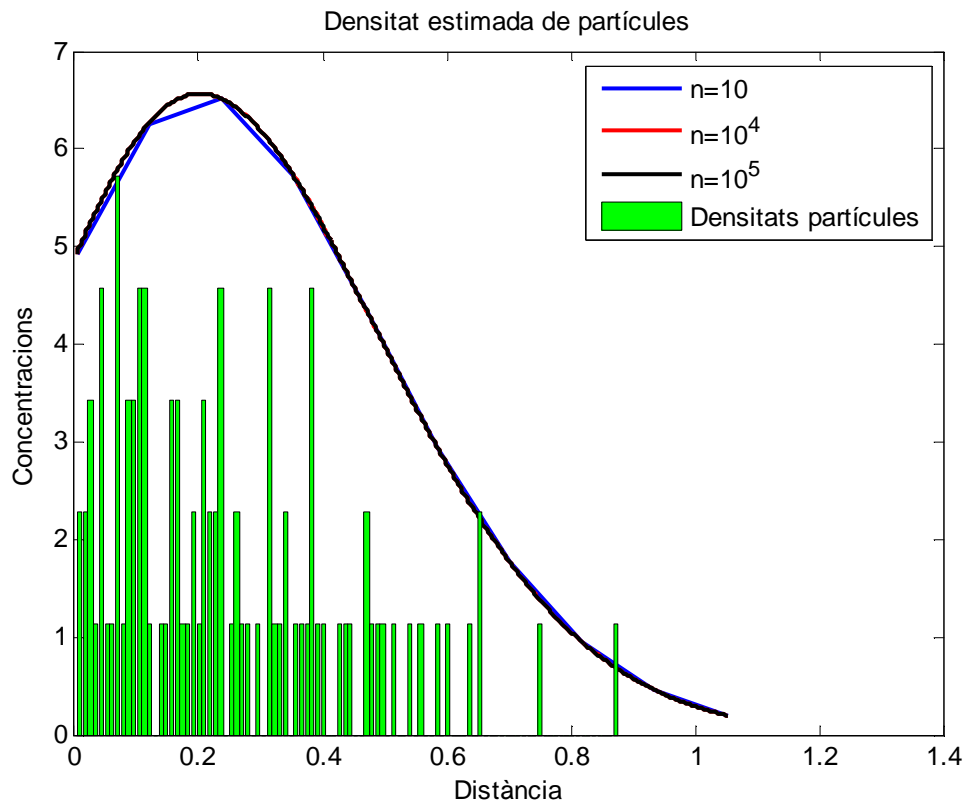


Figura 6.7. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució Beta ($A=1$, $B=3$). (Extreta de MATLAB)

En la distribució Beta, veiem la primera distribució asimètrica com a tal, on no s'observen canvis significatius respecte el número de punts, això és degut a la gran quantitat de punts on avaluem aquesta funció.

6.3.2 Distribució Beta: Bandwidth

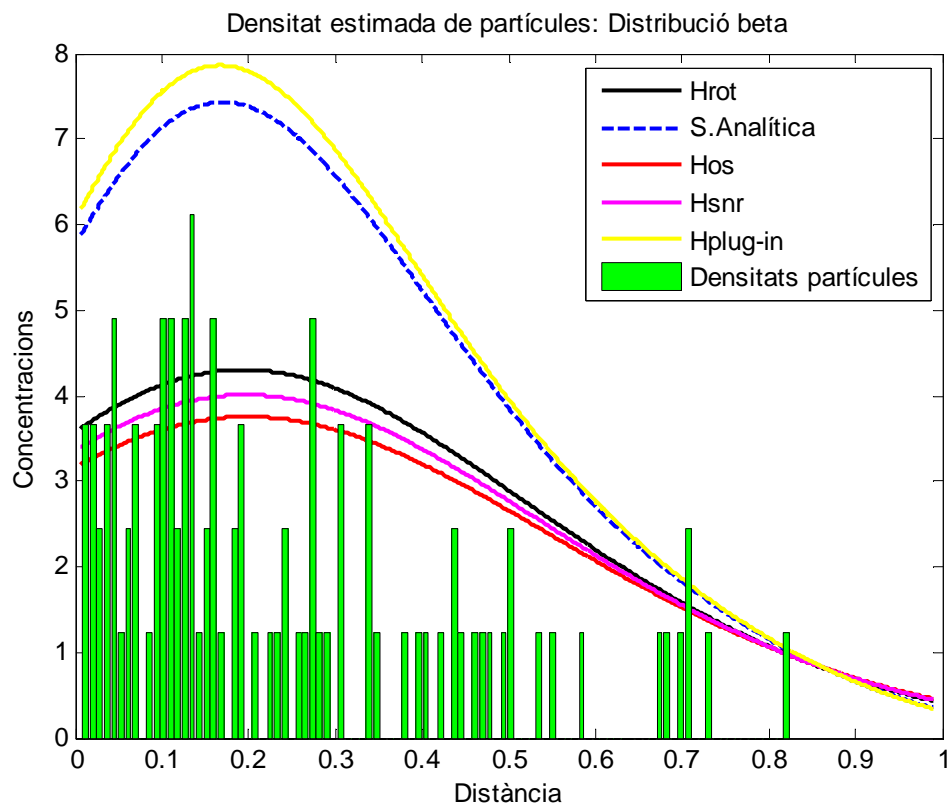


Figura 6.8. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució beta ($A=1$, $B=3$). (Extreta de MATLAB).

La diferència entre els mètodes de *rule of thumb* i el *plug-in method* és bastant significativa, on es veuen les carències dels *rule of thumb* i la bona aproximació del *plug-in method*.

6.3.3 Distribució Beta: Error

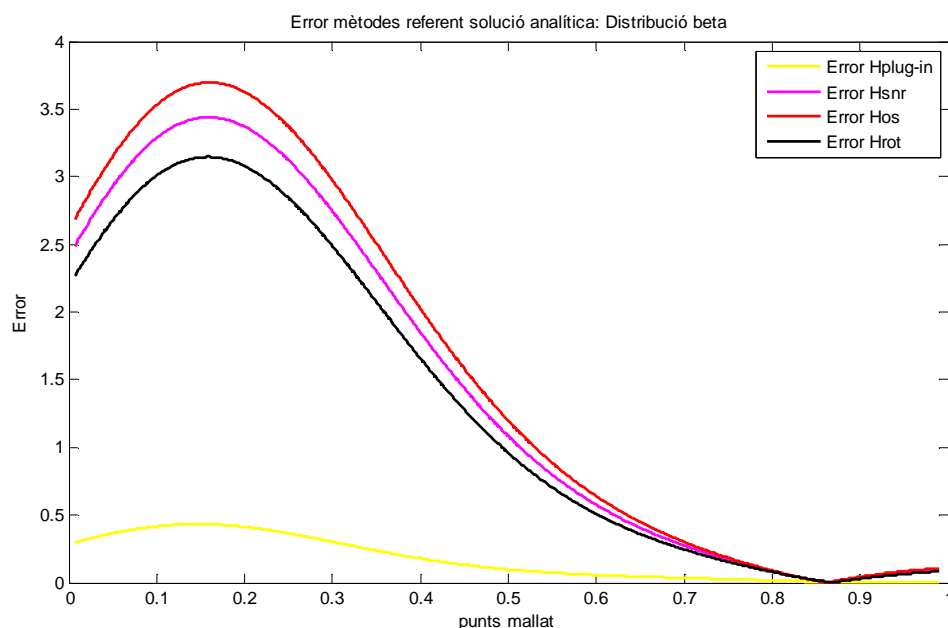


Figura 6.9. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució Beta ($A=1$, $B=3$), (Extreta de MATLAB).

Un cop més, veiem la qualitat del *plug-in method*, respecte la resta de models.

6.4 Distribució Bimodal

La distribució bimodal és aquella que presenta dos grans pics o en aquest cas altes densitats de contaminants, la qual és la seva característica principal. Ens permet avaluar dos tipus de contaminants o arribades del mateix contaminant amb un interval de temps entre les mesures.

6.4.1 Distribució Bimodal: Número de punts

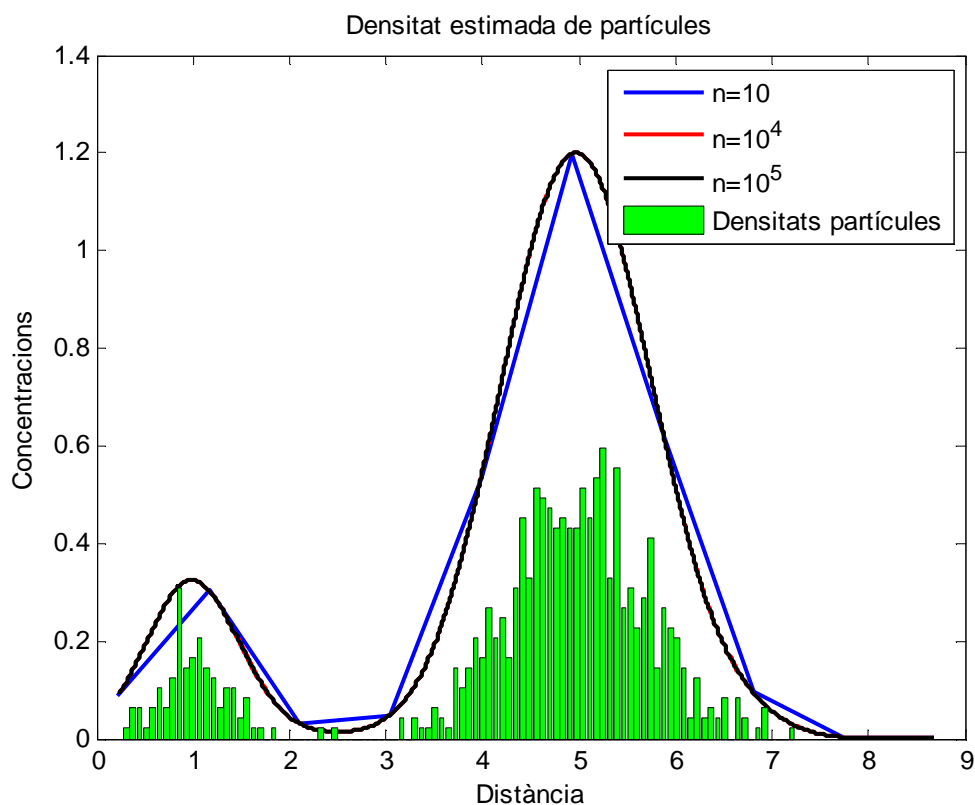


Figura 6.10. Gràfica comparativa segons el número de punts de la densitat de partícules aleatòries amb una distribució Bimodal. (Extreta de MATLAB)

A *Figura 6.10*, observem una diferència més gran amb les funcions de densitats amb un número de punts del mallat menor, que les altres dues funcions (vermella i negra). Això es degut a que la mostra de punts és major que en els altres casos, tenint que representar més punts de mesures reals sobre menys punts de mallat.

6.4.2 Distribució Bimodal: Bandwidth.

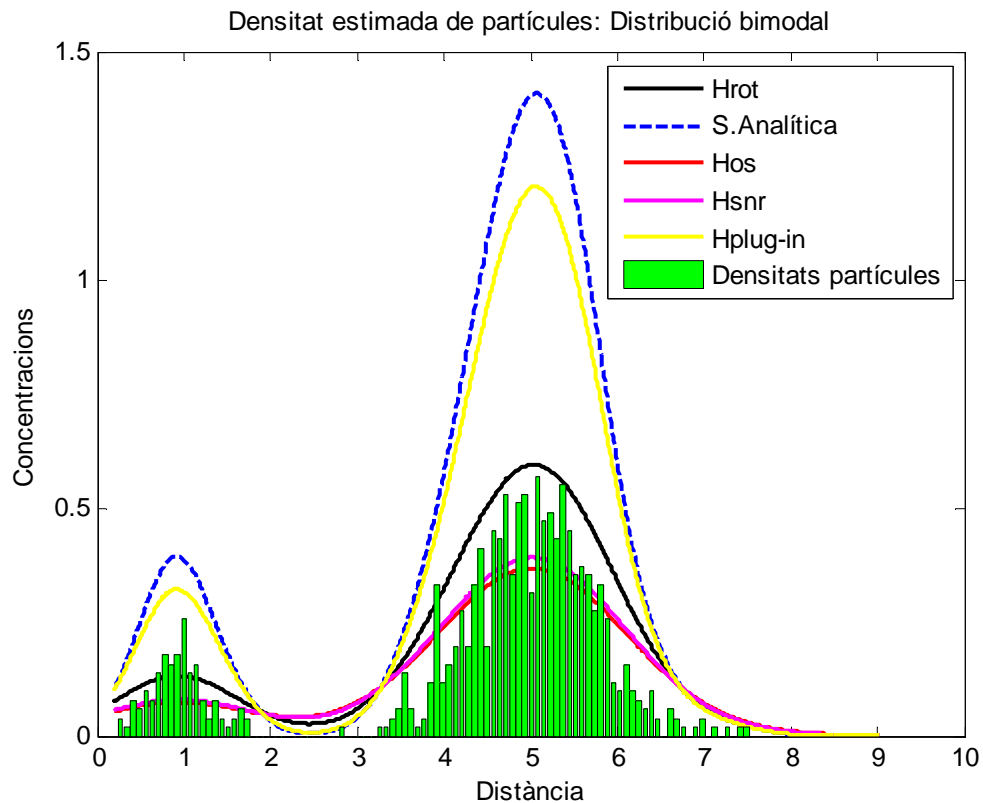


Figura 6.11. Gràfica comparativa segons el coeficient de bandwidth de la densitat de partícules aleatòries en una distribució bimodal. (Extreta de MATLAB).

En aquest darrer mètode s'observa com l'aproximació dels mètodes no és gaire bona degut a les dues famílies de distribucions. No obstant, veiem la clara diferència entre mètodes *rule of thumb* i *plug-in* una vegada més.

6.4.3 Distribució Bimodal: Error

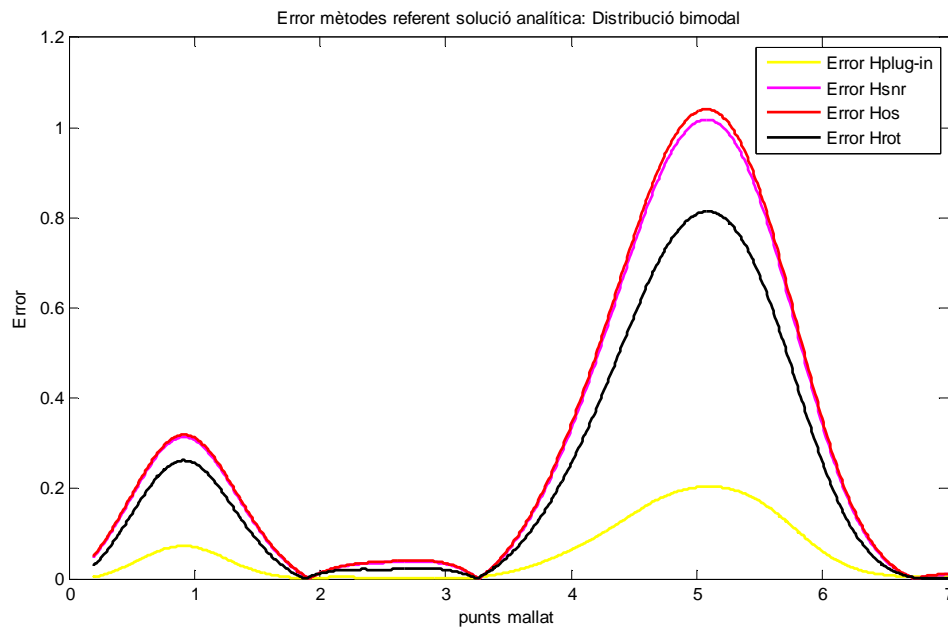


Figura 6.12. Gràfica comparativa dels errors de cada mètode bandwidth respecte la solució analítica en una distribució bimodal, (Extreta de MATLAB).

La diferència d'errors entre mètodes és en alguns, de quasi bé la unitat, tot i que com hem anat veient en totes les distribucions el *plug-in method*, és el mètode més aproximant a la distribució real.

Capítol 7:

Conclusions

7. Conclusions

A partir del estudi que s'ha realitzat, s'ha pogut validar els KDE com a mètode òptim de reconstrucció de densitats de partícules en diferents distribucions aleatòries. És conegut que la distribució de partícules de contaminants en un medi subterrani és complexa i no sempre s'apropa a distribucions conegudes. No obstant s'han estudiat quatre casos per mostrar la funcionalitat del mètode en diferents distribucions. S'ha observat que l'elecció de coeficients és si més no, el pas més important en l'estudi, ja que determina l'aproximació a la realitat de manera eficient.

Com a comportament general, el mallat característic a escollir és important obtenir un número significatiu de punts d'avaluació. Un número escàs de punts obtindrem falses densitats i amb un gran número de punts, tindríem un cost computacional molt alt.

Pel que fa als mètodes de càlcul del BW com els *Rule of Thumb*, donen distribucions molt suavitzades les quals ens serveixen per referenciar una distribució en l'espai, però no per apropar-nos a la densitat real tant com ens agradaria. Amb el *Solve the Equation Plug-in Approach* veiem una òptima aproximació en distribucions simètriques com la distribució normal, i una bona aproximació en distribucions asimètriques com son les distribucions lognormal, beta i bimodal.

La combinació d'un mallat suficient i un bon BW determina l'èxit que obtenim en la reconstrucció de densitats de partícules i degut a l'aleatorietat utilitzada en cada mètode és te un bon registre de casos que garanteixen la funcionalitat del mètode.

S'ha de remarcar les condicions teòriques en les que es desenvolupa aquest estudi, on si que observem una funcionalitat i uns bons resultats, però no oblidem que ens hem centrat en la part final d'un procés molt més ampli que el propi tractament de les dades obtingudes. El coneixement previ del medi, les condicions en el que es troba, el tipus de contaminant que podem trobar-ne, són factor previs a una recollida de dades a tenir en compte. La pròpia recollida de dades és una part molt important, ja que són la eina clau pel desenvolupament de l'estudi, on una presa de mesures sistematitzada, acurada i eficient ens proporciona una situació molt rica amb la que treballar i amb la que obtenir uns resultats vàlids.

Bibliografia

Beardah, C. C., Baxter, M.J. (1996). *MATLAB Routines for kernel Density Estimation and Graphical representation of Archaeological data*. https://www.researchgate.net/publication/2254632_MATLAB_Routines_for_Kernel_Density_Estimation_and_the_Graphical_Representation_of_Archaeological_Data_CC_Beardah . [Últim accés: Agost 2016].]

Crisan, D., Míguez, J. (2014). *Particle-kernel estimation of the filter density in state-space models*. Bernoulli. Vol. 20, núm 4, pp 1879-1929.

Engel, J. et al. (1994). *An iterative bandwidth selector for kernel estimation of densities and their derivatives*. Journaltitle of Nonparametric Statistics. Vol 4, núm 1, pp 21-34.

Fernàndez-Farcia, D., Sanchez-Vila, X. (2011). *Optimal reconstruction of concentrations, gradients and reaction rates from particle distributions*. Journal of Contaminant Hydrology. Vol. 120-121, pp 99-14.

Galeano, C. H. (2009). *Técnicas de solución numérica de la ecuación de Difusión-Advección-reacción para el estudio de dispersión de contaminantes*. Tesis de Doctorado. Facultad de Ciencias. Universidad nacional de Colombia. Bogotá.

Jones, M. C. et al. (1996). *A Brief Survey of Bandwidth Selection for Density Estimation*. Journal of the American Statistical Association. Vol. 91, issue 433, pp 401-407.

Klemelä, J. (2000). *Estimation of Densities and Derivatives of Densities with Directional Data*. Journal of multivariate Analysis. Vol. 73, pp 18-40.

Oyarzún, R. (2007). *Trasporte de contaminantes en aguas subterráneas*. http://www.aulados.net/Temas_ambientales/Contaminantes_aguas_subterraneas/Trasporte_contaminantes.pdf . [Últim accés: Agost 2016].

Pedretti, D., Fernàndez-Garcia, D. (2013). *An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves from particle distributions*. Advances in Water Resources. Vol 59, pp 52-65.

Rivera, A. (2008). *Groundwater Sustainable Development in Canada – Emergin Issues*. Journal of the Geological Association of Canada. Vol. 35, núm 2.

Salamon, P. et al. (2006). *A review and numerical assessment of the random walk particle tracking method*. Journal of Contaminant Hydrology. Vol 87, pp 277-305.

Salamon, P. et al. (2006). *Modeling mass transfer processes using random walk particle tracking*. Water Resources Research. Vol. 42, issue 11.

Sánchez, F. J. (2014). *Ley de Darcy. Conductividad hidráulica*. http://hidrologia.usal.es/temas/Ley_Darcy.pdf. [Últim accés: Agost 2016].

Sheater, S. J. (2004). *Density Estimation*. Statistical Science. Vol 19, núm 4, pp 588-597.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Zheng, C., Bennett, G. D. (2002). *Applied Contaminant Transport Modeling Second Edition*. Wiley Interscience.

